



Analysis on Airbnb Price Determinants in Europe

Presenters : Uma Maheswara R Meleti, Hemanth Vasireddy, Trinath Sai Subhash Reddy Pittala

Math 8050 Data Analysis



Outline

- Problem Statement
- Dataset
- Exploratory Data Analysis
- Outlier Analysis
- Methods
- Model Diagnostics
- Modelling Results
- Conclusion
- References



Problem Statement

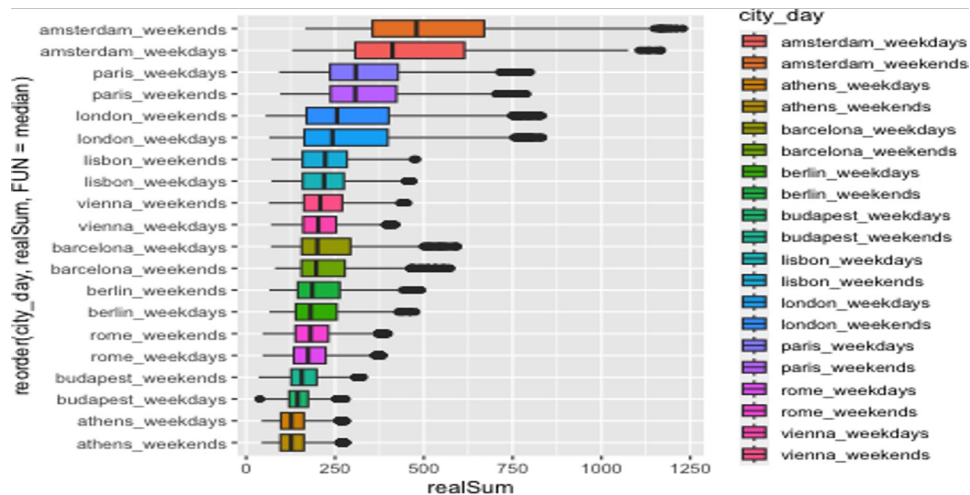
- Understanding the factors that determine Airbnb prices is essential for both hosts and guests.
- This report explores the determinants of Airbnb prices in Europe using data from the publicly available dataset 'Airbnb Price Determinants' from [Source](#).
- In this study, we have explored linear and polynomial regression and random forest models to determine the price of a room and found the main determinants based on the coefficient values.



Dataset

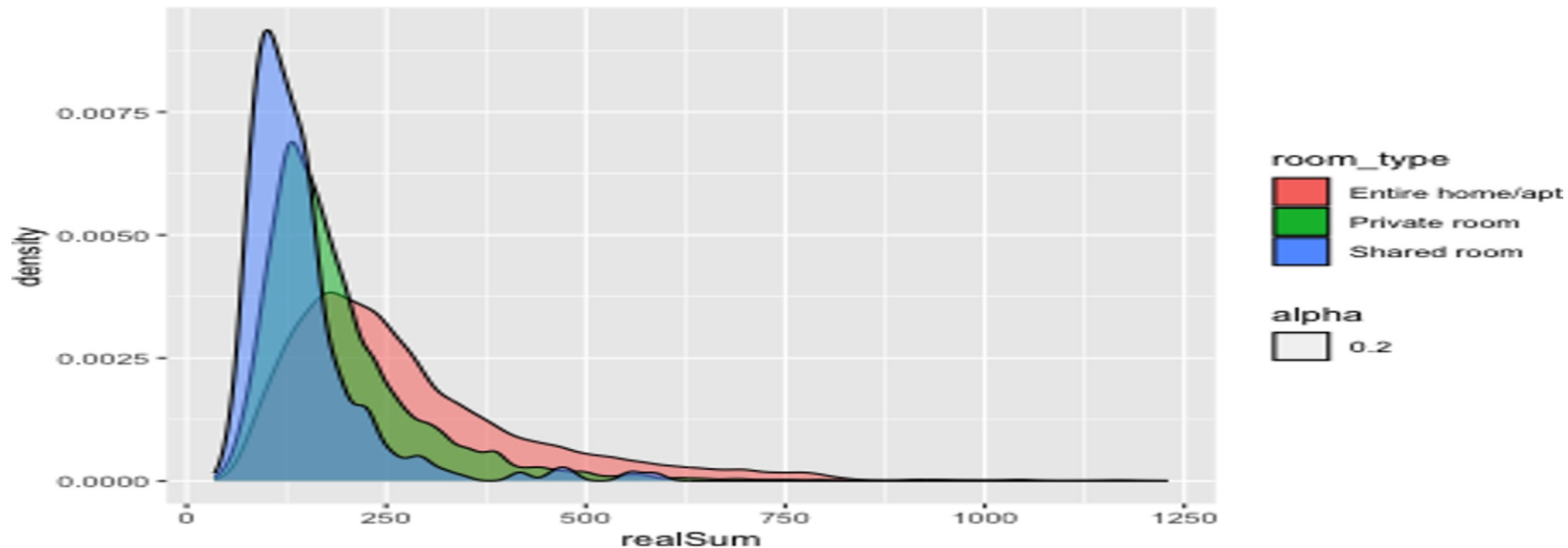
- The dataset consists of data from 10 cities in Europe in 20 files with both weekdays and weekends.
- The dataset consists of
 - Continuous variables : realSum, dist, metro_dist, lat, lng, attr_index, attr_index_norm, rest_index, rest_index_norm
 - Ordinal : person_capacity, guest_satisfaction_overall, cleanliness_rating, bedrooms
 - Nominal : room_type, room_shared, host_is_superhost, multi, biz

Results From EDA



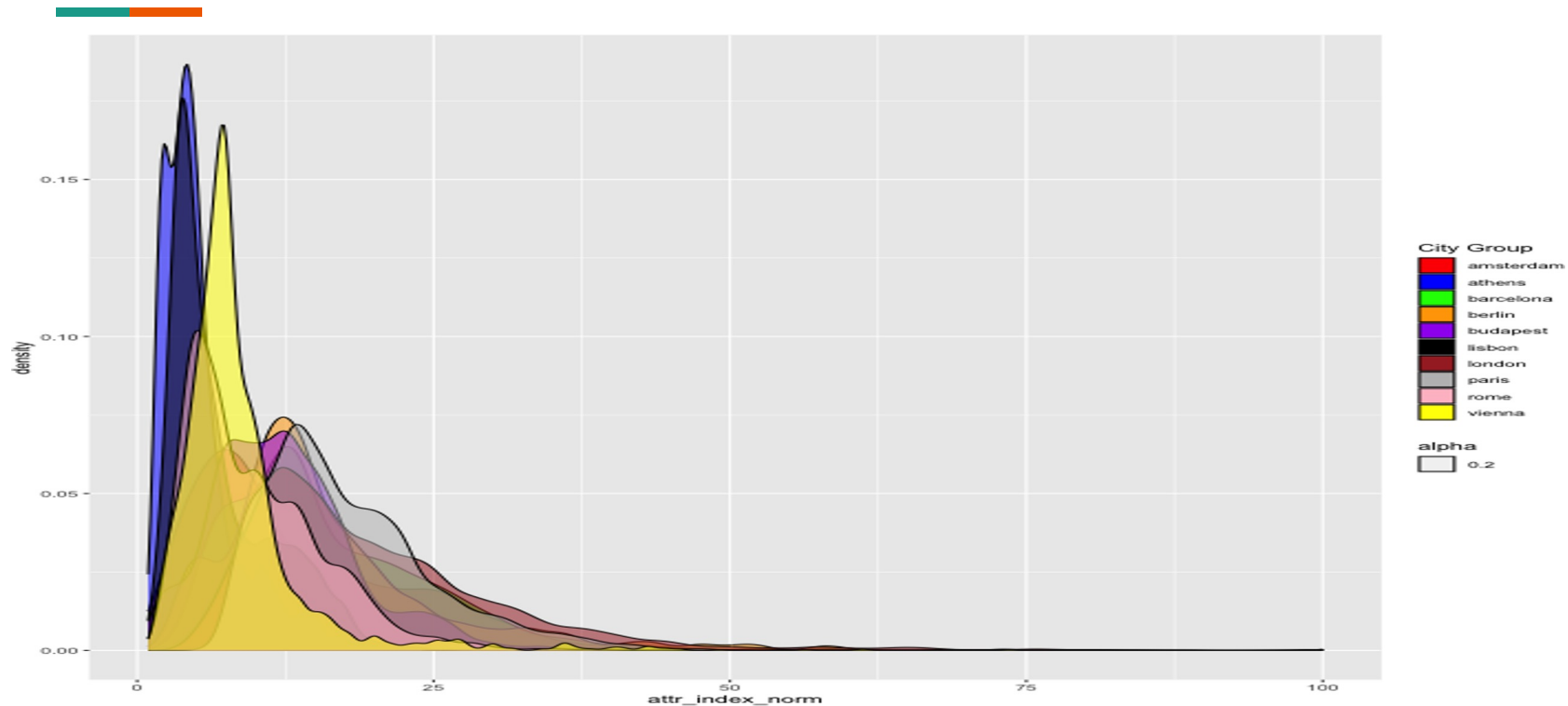
The priciest hotels in Europe are found in Amsterdam.

Results from EDA



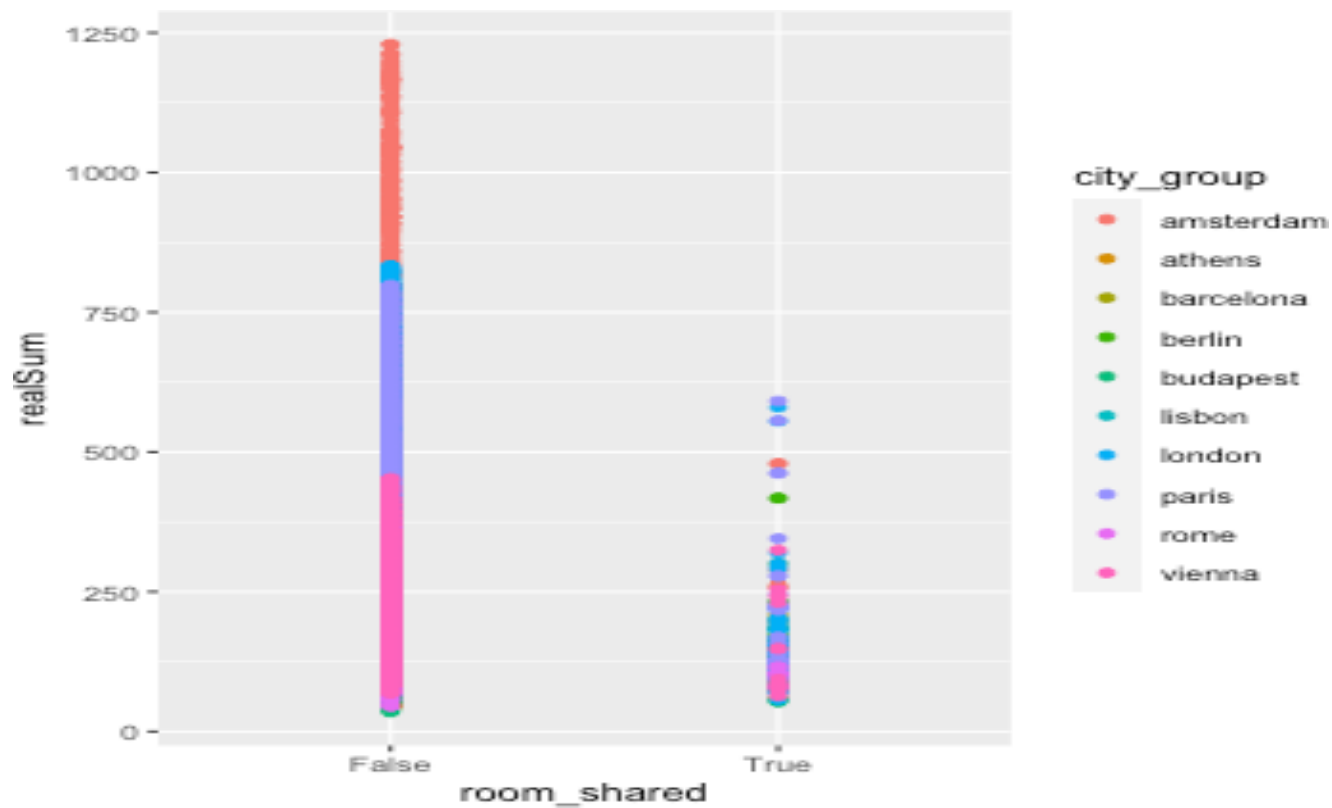
It's Costlier to rent an entire house/apartment compared to other types of rooms.

Results from EDA



The attractive index of cities is varying across different cities it is higher for some cities like Paris.

Results from EDA





Methods Used

- Linear regression
- Polynomial regression
- Interaction Variables
- Stepwise Regression
- Lasso
- Random Forest Regression



Modelling Results

Model	var_names	var_coefs	Model	var_names	var_coefs	Model	var_names	var_coefs	Model	var_names	var_coefs	Model	var_names	var_coefs
M_0	(Intercept)	453.805963	M_1	(Intercept)	677.7774655	M_2	(Intercept)	81087.31478	M_3	(Intercept)	3313.326331	M_4	(Intercept)	14214.11864
M_0	room_typePrivate room	384.5953101	M_1	room_typePrivate room	95.38419868	M_2	room_typePrivate room	148.9417279	M_3	room_typePrivate room	83.28285599	M_4	room_typePrivate room	25.47757592
M_0	room_typeShared room	136.0046307	M_1	room_typeShared room	42.51664607	M_2	room_typeShared room	78.35566832	M_3	room_typeShared room	32.55688542	M_4	room_typeShared room	12.36521125
M_0	person_capacity	119.9608563	M_1	person_capacity	31.890407	M_2	person_capacity	38.86525794	M_3	person_capacity	32.36241001	M_4	person_capacity	6.824991537
M_0	host_is_superhostTrue	42.08786555	M_1	host_is_superhostTrue	16.91133475	M_2	host_is_superhostTrue	18.18910642	M_3	host_is_superhostTrue	20.77860946	M_4	host_is_superhostTrue	5.296527081
M_0	multi	20.6527087	M_1	multi	4.421357529	M_2	multi	12.07097956	M_3	multi	12.77790391	M_4	multi	1.652498502
M_0	biz	16.61275571	M_1	biz	2.725004443	M_2	biz	6.985320078	M_3	biz	6.194671034	M_4	biz	0.462524562
M_0	cleanliness_rating	3.757397212	M_1	cleanliness_rating	2.640302855	M_2	cleanliness_rating	2.492178554	M_3	cleanliness_rating	4.446168901	M_4	cleanliness_rating	-0.492631965
M_0	guest_satisfaction_overall	2.738196195	M_1	guest_satisfaction_overall	2.335020293	M_2	guest_satisfaction_overall	1.774791262	M_3	guest_satisfaction_overall	1.658915477	M_4	guest_satisfaction_overall	-2.915525964
M_0	bedrooms	2.025200616	M_1	bedrooms	1.221665363	M_2	bedrooms	0.215433391	M_3	bedrooms	0.831750876	M_4	bedrooms	-13.33265099
M_0	dist	-2.83720865	M_1	dist	1.053387666	M_2	dist	0.055349136	M_3	dist	0.423821302	M_4	dist	-13.38768559
M_0	metro_dist	-7.074399447	M_1	metro_dist	-23.4210691	M_2	metro_dist	-9.749578513	M_3	metro_dist	0.346180902	M_4	metro_dist	-17.02253017
M_0	attr_index_norm	-21.61855664	M_1	attr_index_norm	-24.66793817	M_2	attr_index_norm	-302.2199647	M_3	attr_index_norm	-39.28949375	M_4	attr_index_norm	-47.92355904
M_0	rest_index_norm	-60.95450514	M_1	rest_index_norm	-29.95296764	M_2	rest_index_norm	-311.0956042	M_3	rest_index_norm	-102.8898274	M_4	rest_index_norm	-48.61561818
M_0	lng	-175.0198029	M_1	lng	-65.86473148	M_2	lng	-384.8001582	M_3	lng	-120.4611811	M_4	lng	-114.5660309
M_0	lat	-340.2468974	M_1	lat	-28023.54948	M_2	lat	-1939.911175	M_3	lat	-221.3241991	M_4	lat	-251.7212566
M_5	(Intercept)	24.49430744	M_6	(Intercept)	6546.714298	M_7	(Intercept)	41416.99566	M_8	(Intercept)	428.4623235	M_9	(Intercept)	408.940146
M_5	room_typePrivate room	20.76733214	M_6	room_typePrivate room	163.4989324	M_7	room_typePrivate room	130.3445018	M_8	room_typePrivate room	44.47609905	M_9	room_typePrivate room	88.51887853
M_5	room_typeShared room	8.685270433	M_6	room_typeShared room	36.38305957	M_7	room_typeShared room	81.79791623	M_8	room_typeShared room	12.32210245	M_9	room_typeShared room	36.04103653
M_5	person_capacity	5.728057336	M_6	person_capacity	25.54307154	M_7	person_capacity	56.22439359	M_8	person_capacity	3.882518636	M_9	person_capacity	20.32350711
M_5	host_is_superhostTrue	4.819142536	M_6	host_is_superhostTrue	12.82925563	M_7	host_is_superhostTrue	43.15533016	M_8	host_is_superhostTrue	3.699383339	M_9	host_is_superhostTrue	17.90834436
M_5	multi	1.238674167	M_6	multi	9.982245594	M_7	multi	25.20968259	M_8	multi	3.695177412	M_9	multi	4.390508977
M_5	biz	0.374695799	M_6	biz	0.933858045	M_7	biz	12.5560253	M_8	biz	1.475050781	M_9	biz	1.60480035
M_5	cleanliness_rating	0.177646654	M_6	cleanliness_rating	0.207399315	M_7	cleanliness_rating	8.107228556	M_8	cleanliness_rating	0.716854353	M_9	cleanliness_rating	1.352556262
M_5	guest_satisfaction_overall	0.130448297	M_6	guest_satisfaction_overall	0.113545147	M_7	guest_satisfaction_overall	7.409335312	M_8	guest_satisfaction_overall	0.513052541	M_9	guest_satisfaction_overall	1.242051249
M_5	bedrooms	-8.337627302	M_6	bedrooms	-19.2800813	M_7	bedrooms	1.641809145	M_8	bedrooms	-0.225693578	M_9	bedrooms	-12.84826556
M_5	dist	-12.56848529	M_6	dist	-21.77143781	M_7	dist	1.3085809	M_8	dist	-0.783530606	M_9	dist	-13.17443006
M_5	metro_dist	-59.19710383	M_6	metro_dist	-22.96385155	M_7	metro_dist	-4.603578145	M_8	metro_dist	-1.508888739	M_9	metro_dist	-13.85744129
M_5	attr_index_norm	-82.58885424	M_6	attr_index_norm	-130.3824975	M_7	attr_index_norm	-97.88170929	M_8	attr_index_norm	-6.252414347	M_9	attr_index_norm	-16.13640861
M_5	rest_index_norm	-190.4190079	M_6	rest_index_norm	-181.4255816	M_7	rest_index_norm	-284.4850304	M_8	rest_index_norm	-39.8036185	M_9	rest_index_norm	-28.48935415
M_5	lng	-474.7694686	M_6	lng	-193.4634265	M_7	lng	-715.552286	M_8	lng	-94.72375311	M_9	lng	-120.2455734
M_5	lat	-2021.541971	M_6	lat	-212.9519776	M_7	lat	-821.2043548	M_8	lat	-17868.39131	M_9	lat	-20955.97861

M_0 - Amsterdam, M_1 - Athens, M_2 - Barcelona, M_3 - Berlin, M_4 - Budapest, M_5 - Lisbon, M_6 - London, M_7 - Paris, M_8 - Rome, M_9 - Vienna

term	estimate	std.error	statistic	p.value
city_dayathens_weekdays	6315.09061077375	1388.53511239174	4.54802370816251	5.43276632399387E-06
city_dayathens_weekends	6303.53106566887	1388.65271393214	4.53931425937278	5.66191686814906E-06
city_daybudapest_weekends	3929.67339172125	706.843976888775	5.55946364432218	2.72512019627685E-08
city_daybudapest_weekdays	3902.35110769949	706.88057311247	5.52052391327863	3.40308057888361E-08
city_dayvienna_weekdays	3231.8185121282	582.709191916445	5.54619449454577	2.93992522226129E-08
city_dayvienna_weekends	3230.26798535313	582.79716507451	5.54269680591213	2.9992285987412E-08
city_dayrome_weekends	2952.53521434873	881.429677236262	3.34971159991624	0.000809784171597337
city_dayrome_weekdays	2947.68523500212	881.398370960292	3.34432798167143	0.00082565958973142
city_dayberlin_weekends	1958.88440687424	342.040071097647	5.72706116154153	1.02993861536248E-08
city_dayberlin_weekdays	1949.04242701527	342.124495503847	5.69688067539542	1.22965025210763E-08
city_daybarcelona_weekends	429.652941610428	837.810922281234	0.512828050081453	0.608074739358428
city_daybarcelona_weekdays	411.771664434536	837.79087513094	0.491496955454641	0.623077988219532
lat	123.211714028539	76.5227515358753	1.61013177853093	0.107377819824515
bedrooms	86.0154191816913	3.18878180193731	26.9743822325609	1.10487900248437E-158
city_dayamsterdam_weekends	67.9410312574828	16.0017245274658	4.24585682254852	2.18302692822801E-05
biz	33.2805902980968	4.18846537329009	7.94577185962373	1.98543351601859E-15
person_capacity	23.962639711116	1.76448699475381	13.580513646381	6.6229089695509E-42
multi	9.60044781984104	4.13239391830793	2.32321700438764	0.0201730153901032
attr_index_norm	6.37045498927847	0.294558501285674	21.6271299639054	4.50013019203766E-103
cleanliness_rating	5.03832622544896	2.41532715274561	2.08598086587221	0.0369873477630624
host_is_superhostTrue	1.07487304905057	3.93436342704325	0.273201260885642	0.784700071060939
guest_satisfaction_overall	0.776010513381954	0.261452270487271	2.96807716351323	0.00299865383859512
rest_index_norm	-0.183747213632232	0.177364361505566	-1.03598723031214	0.300215028045804
dist	-1.53304759323782	1.26282244844453	-1.21398506585478	0.224761355845419
metro_dist	-3.99666546744126	2.50252382009665	-1.59705391626878	0.110262427719338
room_typePrivate room	-114.36546946963	4.28228402903422	-26.7066520329393	1.28065674381786E-155
room_typeShared room	-204.184154195175	18.9348057155667	-10.7835357416586	4.52732676348359E-27
lng	-262.890929428664	4.930553101859	-6.54070528850891	6.20447859714726E-11
city_dayparis_weekdays	-403.028852808039	278.481924100759	-1.44723523478032	0.147839720417429
city_dayparis_weekends	-422.138907209481	278.643708051156	-1.51497735284222	0.129786880370265
city_daylondon_weekdays	-1409.29965386078	206.104550214087	-6.83779010408502	8.17001208243907E-12
city_daylondon_weekends	-1410.93275121758	206.122303369817	-6.84512412364295	7.76270471372057E-12
city_daylisbon_weekends	-2304.00669656451	1143.81264744771	-2.0143217525227	0.0439831538834079
city_daylisbon_weekdays	-2312.9309439741	1143.89771522079	-2.02197356738986	0.0431864319357264
(Intercept)	-4954.90756116218	3996.18236913533	-1.2399102702198	0.215016631100909

coefficients of each model arranged in descending order. The coefficient with a larger value is the essential determinant of the hotel room price.

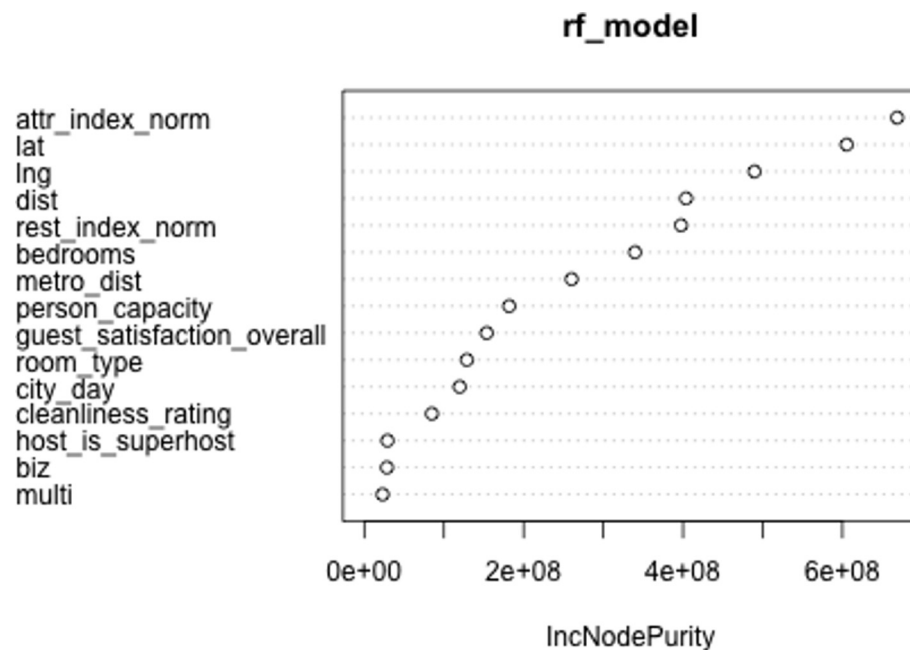
According to the table, all the models have the same descending order of coefficients. The order is as follows:

1 - room_type 2 - person_capacity 3 - host_is_superhost 4 - multi 5 - biz 6 - cleanliness_rating

7 - guest_satisfaction_overall 8 - bedrooms 9 - dist 10 - metro_dist 11 - attr_index_norm 12 -

rest_index_norm, 13 - lng 14 - lat

Random Forest





Model	MLR	MLR with Step	MLR with IVs	MLR with IVs, Step	Poly with Order 2	Poly with Order 2, Step	Poly with Order 2, IVs	Poly with Order 2, IVs, Step
Train R^2	0.2150414	0.2149964	0.2159815	0.2159435	0.2154699	0.2154289	0.22214	0.2221344
Train Adj R^2	0.2146725	0.2146275	0.215613	0.215575	0.2151012	0.2150602	0.2217744	0.2217689
Train $RMSE$	304.9175	304.9262	304.7348	304.7422	304.8342	304.8422	303.5356	303.5367
Test R^2	0.33744	0.3373441	0.3379444	0.3380392	0.3373538	0.3373204	0.334993	0.3350694
Test Adj R^2	0.3371287	0.3370327	0.3376333	0.3377281	0.3370424	0.337009	0.3346805	0.3347569
Test $RMSE$	233.2618	233.2787	233.173	233.1563	233.277	233.2829	233.6922	233.6788
Model	Poly with Order 3	Poly with Order 3, Step	Poly with Order 3, IVs	Poly with Order 3, IVs, Step	Lasso With Order 1	Lasso With Order 2	Lasso With Order 3	Random Forest
Train R^2	0.2160624	0.2160337	0.2330663	0.2285384	0.215693	0.2201963	0.2228371	0.8747755
Train Adj R^2	0.215694	0.2156653	0.2327059	0.2281759	0.2147166	0.217819	0.2166688	0.8724317
Train $RMSE$	304.7191	304.7247	301.3962	302.2846	325.7776	326.6683	326.2303	121.7878
Test R^2	0.3375855	0.3376519	0.1901115	0.3281442	0.3372306	0.1830879	0.02854527	0.7588743
Test Adj R^2	0.3372742	0.3373407	0.189731	0.3281442	0.3353023	0.1772535	0.01036279	0.7543612
Test $RMSE$	233.2362	233.2245	257.8955	234.8925	233.2987	259.0113	282.4505	140.719

Discussion



- When we applied separate models for each city data, on linear regression, latitude was least ranked, but when we combined all the data latitude is on top rank next to city day, it indicates the information about location is being captured by the latitude when all data is combined, in separate models all the data about city represents its location so latitude is least ranked.
- The attraction index is ranked higher in all cities combined model than separated models.
- The host_is_super host is ranked high in individual models than combined model, but its significance value is very low.
- Room Type is ranked at the top in separate models, but in the combined model is at quite a low rank. It is due to the information of Room Type capturing the information about the location;



Conclusion

- Of all the models we have tried, Random Forest Regression performed best with a adjusted R^2 value of 0.7543
- and RMSE of 140.719 compared to the best regression model, which has Adjusted R^2 value of 0.3377 and
- RMSE of 233.1563 on the test set.
- The top ranking of coefficients(Price Determinants) according to the Random Forest model are Attraction index, Location, Distance from the city centre and the distance to Metro station.



References

[Kaggle Data Source](#)

[Determinants of Airbnb prices in European cities: A spatial econometrics approach](#)